Volume 14 Number 4 1986 Nucleic Acids Research

Nucleotide sequence of the hygromycin B phosphotransferase gene from *Streptomyces hygroscopicus*

M.Zalacain, A.González, M.C.Guerrero, R.J.Mattaliano*, F.Malpartida+ and A.Jiménez§

Centro de Biología Molecular, UAM and CSIC, Universidad Autónoma de Madrid, Canto Blanco 28049 Madrid, Spain and *Biogen Research Corp., Cambridge, MA, USA

Received 6 January 1986; Accepted 30 January 1986

ABSTRACT

The nucleotide sequence of a 1467 bp fragment of Streptomyces hygroscopicus DNA containing the gene (hyg) encoding a hygromycin B phosphotransferase (HPH) has been determined. The N-terminal amino acid sequence of HPH determined by automated Edman degradation has allowed the coding sequence of the hyg gene to be identified. The translation initiation triplet is GTG and 5 bp preceding it there is a sequence complementary to the 3'-end of 16S rRNA from S. lividans. The transcriptional start and termination sites have been determined; the presumptive promoter region has only partial homology to that of the Streptomyces vinaceus vph gene and is different to the promoter sequences of other Streptomyces genes.

INTRODUCTION

Streptomyces sp. are mycelial Gram-positive bacteria which produce a large number of compounds with useful clinical, veterinary and agricultural applications. Moreover, these organisms posess a complex life cycle which includes a process of development culminating in spore formation (1). In many instances, antibiotic production by Streptomyces starts when the cultures enter the stationary phase of growth. Therefore, it may be reasoned that both sporulation and antibiotic biosynthesis are regulated by the timely expression of a variety of genes. The relationships between both processes, if any, are not known. Studies of the characterization of the sequences of the genes involved in antibiotic production, would allow the identification of the regions for their transcriptional and translational controls.

Recently, the sequences of several Streptomyces genes, including the aminoglycoside phosphotransferase (aph) gene from S. fradiae (2), the endo-ß-N-acetylglucosaminidase H gene from S. plicatus (3), the rRNA methylase (tsr) gene from S. azureus,

and the viomycin phosphotransferase (vph) gene from S. vinaceus (4) have been determined. The promoter regions of the aph, tsr and vph genes have been characterized (4) as well as those from several other Streptomyces genes (5). It has been shown that Streptomyces promoters contain a much higher T+A content (about 50%) than the corresponding coding sequences (about 27%) and contain several classes of sequences (4) (5). The finding that apparently most of these promoters do not function in Escherichia coli and that a number of E. coli transcriptional signals can be utilized by Streptomyces RNA polymerase(s) (6) (7) indicates that there may be important differences between the controlling regions of the genes from these organisms. The characterization of Streptomyces genes is, therefore, important in order to understand the regulation of their expression. In addition, these studies could throw light about potential pathways of evolution of antibiotic resistance mechanisms (8) (2)

We have cloned and expressed in S. lividans the hyg gene determining a hygromycin B phosphotransferase enzyme of S. hygroscopicus (9). This gene has provided a useful dominant selective determinant for Streptomyces cloning vectors (9) (10). In the present work we have determined the sequence of the hyg gene as well as its sites for initiation and termination of transcription.

MATERIALS AND METHODS

Bacterial strains, plasmids and growth conditions

S.lividans 1326 was provided by Prof. D.A. Hopwood. S.lividans FM4, containing plasmid pFM4, has been described previously (9). E. coli JM83 (11) was provided by Prof. J.E. Davies. Streptomyces plasmid pIJ41 (12) and E. coli plasmid pUC9 (11) were provided by Prof. D.A. Hopwood and Prof. J.E. Davies, respectively.

Streptomyces strains were grown in liquid medium YEME (13) supplemented with 34% sucrose and 5 mM $MgCl_2$. Solid medium R2YE and overlaying agar were described previously (14) (15). Liquid LB, LB agar medium and X-gal plates for E. coli were prepared as described elsewhere (16).

DNA isolation and cloning procedures

Plasmid DNA from Streptomyces and E. coli was prepared

according to (17) and (18), respectively. Plasmid DNA was digested with restriction endonucleases as recommended by the suppliers (Boehringer-Mannheim, New England Biolabs and BRL). DNA fragments were analysed by agarose gel electrophoresis using appropriate molecular weight markers. Specific DNA fragments were isolated from restricted samples of plasmid DNA by gel electrophoresis in low melting-point agarose followed by extraction of the relevant band by the cetyltrimethylammoniumbromide-assisted method (19).

Restricted plasmid DNA was usually treated with calf intestinal alkaline phosphatase (Molecular Biology Grade, Boehringer-Mannheim) to avoid religation. Ligation of DNA fragments was achieved by T_4 DNA ligase (New England Biolabs).

Transformation of *Streptomyces* protoplasts was performed as described elsewhere (12). Protoplasts were regenerated on R2YE medium. After incubation for 8-19 hr at 30°C, 3 ml of soft agar containing either thiostrepton (25 μg/ml) and/or hygromycin B (200 μg/ml) were overlayed. Incubation was then continued for several days.

Transformation of *E. coli* JM83 was carried out as described elsewhere (20). White colonies were selected on LB plates containing ampicillin (100 μg/ml) and X-gal (40 μg/ml).

DNA sequencing

The Maxam and Gilbert chemical degradation method (21) for DNA sequencing was followed throughout this work. 5'-end-labelling was carried out using T_4 polynucleotide kinase (Boehringer-Mannheim) and (γ -^{32}P) ATP (Amersham, 3000 Ci/mmol). 3'-end-labelling was achieved with the Klenow fragment of *E. coli* DNA polymerase I (Boehringer-Mannheim) and (α -^{32}P) dNTP (Amersham, 400 Ci/mmol). After digestion with a second restriction enzyme, end-labelled fragments were separated by low melting-point agarose gel electrophoresis (instead of PAGE), visualized by ethidium bromide staining and purified by the cetyltrimethylammoniumbromide-assisted method (19). If required, in order to obtain overlapping sequence measurements, the size of a restriction DNA fragment was reduced by treatment with *Bal*31 exonuclease (20). The resulting fragments were treated with Klenow polymerase for blunt-ending and subcloned into *Hin*dII-restricted plasmid pUC9.

A compression of C's was solved by a method developed elsewhere (22). After performing the chemical degradation sequencing reactions (21) samples were treated with 1.5 M metoxyamine and 1 M sodium bisulphyte to destroy the cytosine rings. Then, reaction products were subjected to gel electrophoresis.

S1 mapping

Determination of the initiation and termination sites of transcription was achieved by the S1 mapping procedure (23). To prepare RNA, S. hygroscopicus cells were grown at 30°C in liquid media to an OD_{660} of 1.0. Mycelia were collected by centrifugation, converted to protoplasts by lysozyme treatment and lysed with 1% SDS, in the presence of heparin (1 mg/ml) as an RNAse inhibitor. After several extractions with phenol-chloroform, RNA was partially purified by repeated precipitations at 3 M ammonium acetate, pH 6.0. S1 mapping was carried out essentially as described but the incubation temperature for the formation of the DNA-RNA hybrids was 60°C, considering the high C+G content of Streptomyces DNA. After S1 digestion, the reaction products were loaded on a 6% acrylamide sequencing gel (21) in parallel to sequencing reactions of the same labelled fragment used for mapping.

Amino acid sequence analysis

The HPH enzyme was highly purified by affinity chromatography (M. Zalacain, J.M. Pardo and A. Jiménez, unpublished results). The enzyme was further prepared for protein microsequence analysis by modification of a procedure for the isolation of polypeptides from polyacrylamide gels (24). Following electrophoresis in a 10% SDS gel, the R_f of HPH was determined after Coomassie Blue R250 staining of sample lanes cut from the sides of the gel. The region containing the HPH polypeptide was excised from the corresponding position in non-stained lanes. Electroelution of the sample from the gel slices was performed in 0.1% SDS, 10 mM $(NH_4)_2CO_3$ at 50 VDC for 16 hours at 23°C. The eluted polypeptide was dialysed exhaustively at 4°C versus 0.1% SDS. For quantitation purposes, an aliquot of the sample was then hydrolyzed with constant boiling HCl in a sealed, evacuated tube at 110°C for 24 hours and dried under vacuum in a Speed-Vac (Savant). The hydrolysate was analyzed on a Beckman 6300 amino acid analyzer

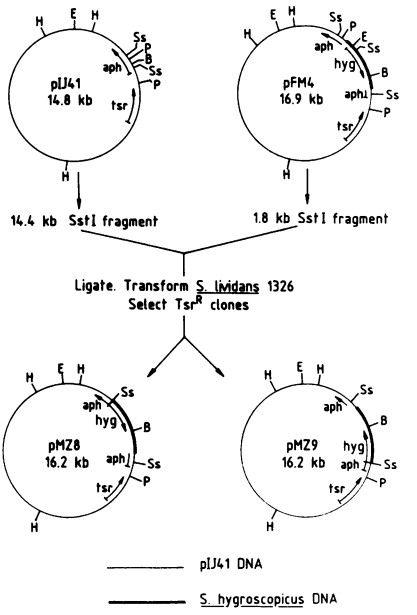

Fig 1. Construction of plasmids pMZ8 and pMZ9. Plasmids pIJ41 and pFM4 were digested with restriction endonuclease SstI, separately. Digestion products were separated by low melting-point agarose (1%) gel electrophoresis and the 14.4 Kb replicon fragment from pIJ41 and the 1.8 Kb fragment from pFM4 were extracted as described in Materials and Methods and then ligated with T_4 ligase. The ligation products were used to transform S. lividans protoplasts. Transformants were selected for resistance to thiostrepton. Clones contained either pMZ8 or pMZ9. The arrows indicate the direction of transcription of the relevant genes. Symbols: aph, gene for aminoglycoside phosphotransferase; tsr, gene for 23S rRNA methylase conferring resistance to thiostrepton; hyg, gene for hygromycin B phosphotransferase.

equipped with ninhydrin detection.

An aliquot of the eluted HPH polypeptide (550 pmoles) was subjected to automated Edman degradation using a gas phase protein sequenator (Applied Biosystems 470A) (25). Identification of phenylthiohydantoin-amino acids form each cycle was performed using two reverse phase HPLC analysis procedures (26).

RESULTS

Identification of the hyg gene

Plasmid pFM4 contains an insertion of a 2.1 Kb MboI fragment of S. hygroscopicus DNA in the BamHI site of pIJ41 (12) (Fig 1)

and expresses hygromycin B phosphotransferase activity when introduced into S. lividans 1326 (9). Insertion of foreign DNA into the single BamHI site of pFM4 does not affect its ability to confer resistance to hygromycin B (10; M. Zalacain, unpublished observations). Conversely, insertion of foreign DNA into the single EcoRI site abolishes the expression of the phosphotransferase activity (M. Zalacain, F. Malpartida and A. Jiménez, unpublished results). Since this enzyme has a MW of about 41000 (27) the hyg gene must be located within the BamHI site and the distal MboI insertion site. To locate the hyg gene more precisely, the 1.8 Kb SstI fragment from pFM4 was introduced into the large SstI fragment from pIJ41. Two types of plasmid were obtained: pMZ8 and pMZ9, differing in the relative orientation of the 1.8 Kb SstI fragment (Fig 1). S. lividans 1326 clones carrying pMZ8 were sensitive to hygromycin B, while those containing pMZ9 were resistant to 200 µg/ml hygromycin B. Cell-free extracts from clones having pMZ8 lacked measurable HPH activity while those from clones carrying pMZ9 had 10-fold lower HPH activity than those obtained from pFM4-containing clones (results not shown). These results suggest that pMZ9 carries either an aph-hyg fused gene or a polycistronic aph-hyg gene construction governed by the aph gene promoter. Sequencing of hyg gene showed that, indeed, the former was the case (see below) and that pMZ9 expresses a fused protein which has low hygromycin B phosphotransferase activity. We conclude that the hyg gene is transcribed from the EcoRI to the BamHI site.

The nucleotide sequence of the hyg gene

Fig 2 shows the restriction endonuclease map of the MboI-

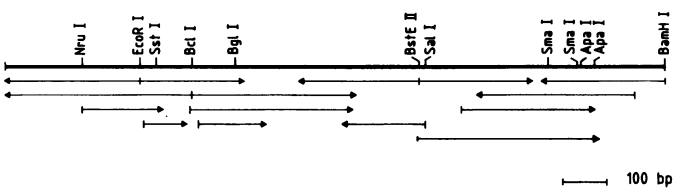

Fig 2. Restriction endonuclease cleavage map and sequencing strategy for the 1467 bp fragment of the S. hygroscopicus DNA. The arrows indicate the direction and length of sequence determined from each labelled end.

BamHI fragment of S. hygroscopicus DNA containing the hyg gene, and the strategy used for its sequencing. Because of the high G+C content (about 70%) in Streptomyces DNA (28), secondary structures may appear in single-stranded chains, resulting in base compressions. This problem happened when sequencing the hyg gene (Fig 3), and could not be eliminated by sequencing the complementary strand. We solved this problem by treatment of the DNA sample, after performing the chemical degradation reactions, with methoxyamine and sodium bisulphite, as described (22), to destroy the cytosine rings and avoid intracatenary complementation. The result clearly showed that the compression was due to the three consecutive C's (Fig 3) present at positions 550-552 of the se-

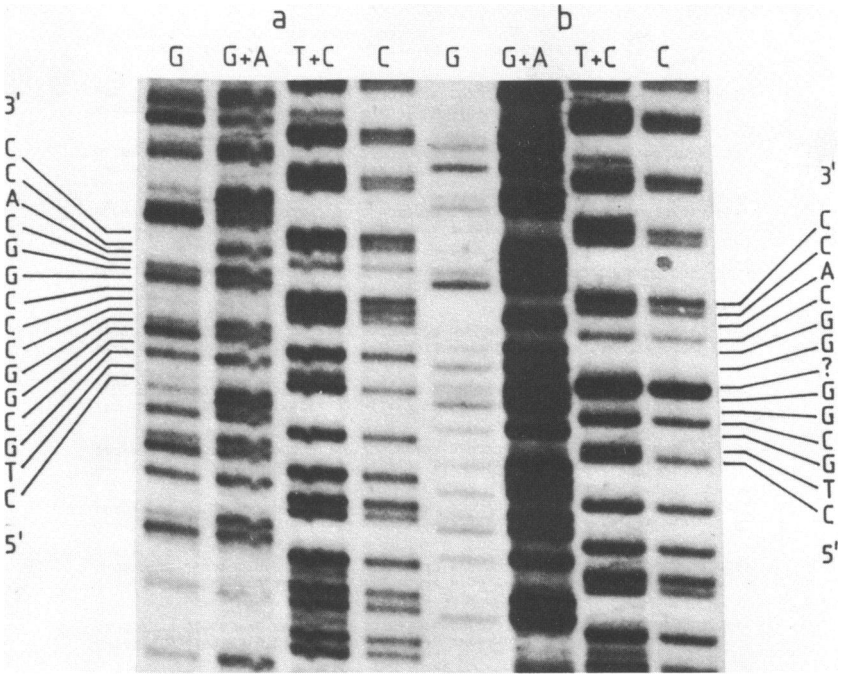

Fig 3. Resolution of the C's compression at positions 550-552. After performing the modification reactions (21), the samples were treated with methoxyamine and sodium bisulphite (22) prior to gel electrophoresis. a) Autoradiography of the gel loaded with samples treated with methoxyamine and sodium bisulphite. Notice the about 10% retardation of the bands as related to the normal sequencing method. b) Autoradiography of the gel loaded with samples treated by the normal sequencing reactions.

```
                MboI                    NarI
   1 GATCGGCGG GGCCTGGCG GCGCCTCGC CGTCGCAGA ACCAGGCGG TGGCGTACA CCGTCGCCT

  64 CGGTCGGCC CGTAGAGAT TGGCGATCC CGACCGCAG CACCACCGA GAACGTCCC CGACGTGGC

                                    ↓
 127 CGACCAGCC CGTCATCGT CAACGCCTG ACCGCGGTG CGGACAGGC CGTGTCGCG ACCGGCCGT

                                              AvaII
 190 GCGGAATTA AGCCGGCCC GTACCCTGT GAATAGAGG TCCGCTGTG ACACAAGAA TCCCTGTTA
                                          Met ThrGlnGlu SerLeuLeu

                                                                 Eco RI
 253 CTTCTCGAC CGTATTGAT TCGGATGAT TCCTACGCG AGCCTGCGG AACGACCAG GAATTCTGG
     LeuLeuAsp ArgIleAsp SerAspAsp SerTyrAla SerLeuArg AsnAspGln GluPheTrp

                                  SstI
 316 GAGCCGCTG GCCCGCCGA GCCCTGGAG GAGCTCGGG CTGCCGGTG CCGCCGGTG CTGCGGGTG
     GluProLeu AlaArgArg AlaLeuGlu GluLeuGly LeuProVal ProProVal LeuArgVal

                                                      BclI
 379 CCCGGCGAG AGCACCAAC CCCGTACTG GTCGGCGAG CCCGACCCG GTGATCAAG CTGTTCGGC
     ProGlyGlu SerThrAsn ProValLeu ValGlyGlu ProAspPro ValIleLys LeuPheGly

 442 GAGCACTGG TGCGGTCCG GAGAGCCTC GCGTCGGAG TCGGAGGCG TACGCGGTC CTGGCGGAC
     GluHisTrp CysGlyPro GluSerLeu AlaSerGlu SerGluAla TyrAlaVal LeuAlaAsp

                          BglI
 505 GCCCCGGTG CCGGTGCCC CGCCTCCTC GGCCGCGGC GAGCTGCGG CCCGGCACC GGAGCCTGG
     AlaProVal ProValPro ArgLeuLeu GlyArgGly GluLeuArg ProGlyThr GlyAlaTrp

 568 CCGTGGCCC TACCTGGTG ATGAGCCGG ATGACCGGC ACCACCTGG CGGTCCGCG ATGGACGGC
     ProTrpPro TyrLeuVal MetSerArg MetThrGly ThrThrTrp ArgSerAla MetAspGly

 631 ACGACCGAC CGGAACGCG CTGCTCGCC CTGGCCCGC GAACTCGGC CGGGTGCTC GGCCGGCTG
     ThrThrAsp ArgAsnAla LeuLeuAla LeuAlaArg GluLeuGly ArgValLeu GlyArgLeu

 694 CACAGGGTG CCGCTGACC GGGAACACC GTGCTCACC CCCCATTCC GAGGTCTTC CCGGAACTG
     HisArgVal ProLeuThr GlyAsnThr ValLeuThr ProHisSer GluValPhe ProGluLeu

 757 CTGCGGGAA CGCCGCGCG GCGACCGTC GAGGACCAC CGCGGGTGG GGCTACCTC TCGCCCCGG
     LeuArgGlu ArgArgAla AlaThrVal GluAspHis ArgGlyTrp GlyTyrLeu SerProArg

 820 CTGCTGGAC CGCCTGGAG GACTGGCTG CCGGACGTG GACACGCTG CTGGCCGGC CGCGAACCC
     LeuLeuAsp ArgLeuGlu AspTrpLeu ProAspVal AspThrLeu LeuAlaGly ArgGluPro

                                                              Bst
 883 CGGTTCGTC CACGGCGAC CTGCACGGG ACCAACATC TTCGTGGAC CTGGCCGCG ACCGAGGTC
     ArgPheVal HisGlyAsp LeuHisGly ThrAsnIle PheValAsp LeuAlaAla ThrGluVal

     EII         SalI
 946 ACCGGGATC GTCGACTTC ACCGACGTC TATGCGGGA GACTCCCGC TACAGCCTG GTGCAACTG
     ThrGlyIle ValAspPhe ThrAspVal TyrAlaGly AspSerArg TyrSerLeu ValGlnLeu

1009 CATCTCAAC GCCTTCCGG GGCGACCGC GAGATCCTG GCCGCGCTG CTCGACGGG GCGCAGTGG
     HisLeuAsn AlaPheArg GlyAspArg GluIleLeu AlaAlaLeu LeuAspGly AlaGlnTrp

1072 AAGCGGACC GAGGACTTC GCCCGCGAA CTGCTCGCC TTCACCTTC CTGCACGAC TTCGAGGTG
     LysArgThr GluAspPhe AlaArgGlu LeuLeuAla PheThrPhe LeuHisAsp PheGluVal

1135 TTCGAGGAG ACCCCGCTG GATCTCTCC GGCTTCACC GATCCGGAG GAACTGGCG CAGTTCCTC
     PheGluGlu ThrProLeu AspLeuSer GlyPheThr AspProGlu GluLeuAla GlnPheLeu

                                                              ↓
1198 TGGGGGCCG CCGGACACC GCCCCCGGC GCCTGACGC CCCGGGCCG CCCGGCGCC GCCCCCGGC
     TrpGlyPro ProAspThr AlaProGly Ala***

1261 CCCCGGCGG CCGCCCGGA GCCCCGCCC GCGCTCGGG AGCCCCGGG CCCGCGCCG AAGCCCGCT

1324 GCTGCGAGC CCGGAGCGG GCCGGCCGA CGGCGGTGC GGGCCCGGC GGCGGACGC TCAGCAGCG

1387 GCGGGCGTG AAAGGCCCT GGCATCCTC GATCATCTC CTCCAGGGT GGTCGGCCG AGCTTCCAT

                    BamHI
1450 CCCAGCTCG GCAAGGATC C
```

quence corresponding to the nonsense strand (Fig 4).

A computerized analysis of the MboI-BamHI sequence presented in Fig 4 showed that none of the three ATG codons at positions 275, 585 and 594 could be the initiation codon for the HPH enzyme, because they correspond to open reading frames that are too small for the HPH enzyme (41000 MW). However, the GTG codon at position 231 was in phase with the TGA triplet at position 1228 (indicated by asterisk, Fig 4). Its derived amino acid sequence, shown in Fig 4, was 37000 MW, close to that of the HPH enzyme determined by gel electrophoresis. Within this coding frame three additional GTG triplets (underlined in Fig 4) are found at positions 358, 367 and 376. To complicate matters: i) there are two ribosomal binding sequences (RBS), complementary to the 3'-end of the S. lividans 16S ribosomal RNA (6), each preceding by 5 and 14 bp the two first GTG codons, respectively (boxed in Fig 4); ii) upstream of the EcoRI site there is a dramatic increase in A+T rich triplets compared to the rest of the sequence; and iii), the sequences starting at the four GTG codons could form a fused gene with the first 120 nucleotides of the aph gene in plasmid pMZ9, maintaining the coding frame for both aph and hyg genes. Such fusions would explain the expression of HPH activity by S. lividans clones containing pMZ9 (Fig 1). Nevertheless, sequences starting at either the second, third or fourth GTG triplets would encode proteins of only 32100, 31800, and 31500 MW, respectively, substantially smaller than that calculated for the HPH enzyme. Therefore, these findings suggest that the first GTG codon is the correct initiation codon for HPH.

To identify the correct initiating triplet of the hyg gene, the $-NH_2$ terminal amino acid sequence of the HPH protein was determined by the Edman-degradation method (24). The results indi-

Fig 4. Nucleotide sequence of the 1467 bp fragment of S. hygroscopicus DNA containing the hyg gene. The two presumptive RBS are boxed. The four potentially initiating codons are underlined. The arrows indicate the two inverted repeat sequences for transcription termination. The vertical arrows indicate the sites of initiation and termination of transcription. The NarI and AvaII sites indicated (nucleotides 18 and 224) are not unique in the sequenced DNA; they are indicated to mark a fragment, containing none more of these sites internally, with presumptive promoter activity.

Table 1. Amino acid sequence of intact 41000 Dalton HPH*.

Cycle	Amino acid	Yield (pmol)	Amino acid	Yield (pmol)
1	THR	67	SER	59
2	GLN	150	LEU	90
3	GLU	147	LEU	104
4	SER	30	LEU	89
5	LEU	167	(LEU)	
6	LEU	142	ASP	22

* amount applied = 550 pmoles.

cated a heterogeneus amino terminal: thr-gln-glu-ser-leu-leu- or ser-leu-leu-leu-leu-asp (Table 1). This result confirms that the first GTG codon is the functional initiating triplet of the hyg gene (Fig 4). It appears that, after translation, the HPH protein suffers the removal of either the initiating methionine residue or the first four amino acids, leaving threonine or serine as the amino-terminal residue, respectively. Alternatively, it may be possible that two reactions take place sequentially.

Codon usage in the hyg gene

As it was found previously with the aph, tsr and vph genes (2) (4) the coding sequence of the hyg gene contains a high proportion (91%) of codons that use a G or C at the 3' position. In this respect, it is curious to note that the sequence starting at the second GTG codon has a higher proportion (94%) of codons containing a G or C at the third position. This is due to the higher A+T content (28%) in this position of the codons within the first and second GTG triplets of the hyg gene.

Table 2 indicates the codon usage in hyg gene. It is clearly non-random and seventeen out of sixty-one coding triplets are not used, and ten are used only once.

Initiation and termination of transcription in the hph gene

S1 nuclease mapping was performed to determine the initiation and termination sites of transcription. The results showed that the transcription start site lies at or around the G 75 bp upstream from the initiator GTG triplet (Fig 5 (A); Fig 4, nu-

Table 2. Codon usage for the hyg gene.

Phe	TTT	0	Ser	TCT	0	Tyr	TAT	1	Cys	TGT	0
	TTC	14		TCC	6		TAC	5		TGC	1
	TTA	1		TCA	0	***	TAA	0	***	TGA	1
	TTG	0		TCG	4	***	TAG	0	Trp	TGG	9
Leu	CTT	1	Pro	CCT	0	His	CAT	2	Arg	CGT	1
	CTC	15		CCC	10		CAC	6		CGC	12
	CTA	0		CCA	0	Gln	CAA	2		CGA	1
	CTG	33		CCG	16		CAG	3		CGG	13
Ile	ATT	1	Thr	ACT	0	Asn	AAT	0	Ser	AGT	0
	ATC	4		ACC	19		AAC	6		AGC	5
	ATA	0		ACA	1	Lys	AAA	0	Arg	AGA	0
Met	ATG	4		ACG	2		AAG	2		AGG	1
Val	GTT	0	Ala	GCT	0	Asp	GAT	5	Gly	GGT	1
	GTC	8		GCC	14		GAC	21		GGC	16
	GTA	1		GCA	0	Glu	GAA	8		GGA	2
	GTG	14		GCG	14		GAG	20		GGG	7

cleotide 157). The termination site lies at or around the G 22 bp downstream from the terminator TGA triplet (Fig 5 (B); Fig 4, nucleotide 1252). Two imperfect inverted repeats are found downstream from the site for termination of transcription (Fig 4). The smaller one could form a hairpin-loop structure with a theoretical free energy of formation of -13.6 Kcal/mol, while this value is -85.2 Kcal/mol (29) for the bigger one.

DISCUSSION

The nucleotide sequence of a 1467 bp DNA fragment encoding a hygromycin B phosphotransferase from Streptomyces hygroscopicus has been determined. In assessing this sequence it was important to clarify an ambiguity due to compression of C residues in gel electrophoresis during sequencing. Treatment of the DNA samples with methoxyamine and sodium bisulphite revealed the correct number of C's at positions 550-552. This technique may be useful in the sequencing of Streptomyces DNA whose high G+C content could result in compressed bands.

The derived sequence contains four potential initiating GTG codons in phase with a TGA termination triplet. However, the

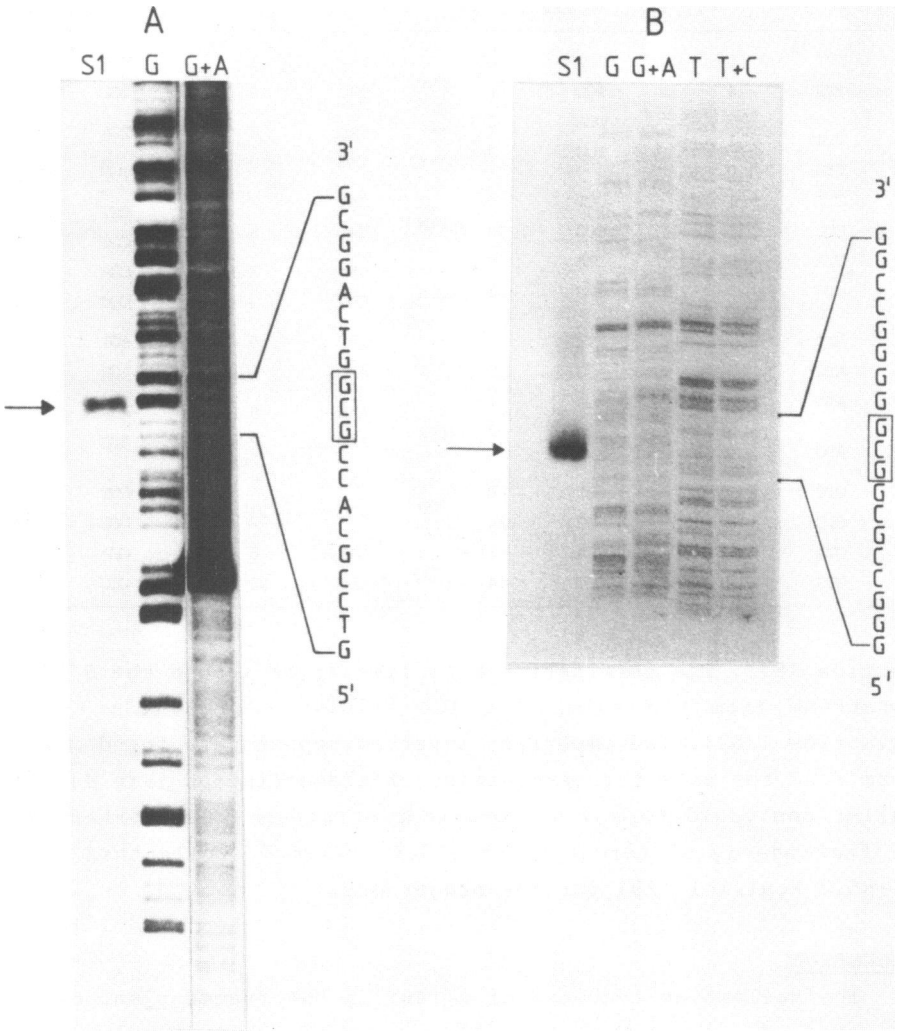

Fig 5. S1 mapping to determine the sites for transcription initiation (A) and termination (B). For details see Materials and Methods.

first GTG, located at position 231, encodes a polypeptide of MW 37000, the closest value to that of about 41000, as determined by physicochemical methods, (27; M. Zalacain unpublished results) for the native protein. Determination of the amino terminal protein sequence of electrophoretically pure HPH enzyme showed that this GTG is the correct initiating codon. The presence of an in-

phase TGA termination codon, four triplets upstream of the proposed initiator codon, eliminates the possibility of a post-translationally processed protein giving thr-gln-glu as the amino terminal end. Protein sequencing results indicate that processing in addition to elimination of the formyl-methionine moiety, is carried out on the HPH protein, resulting in the elimination of the terminal thr-gln-glu tripeptide. A similar heterogeneous NH_2-terminal sequence was found for the endo-ß-N-acetylglucosaminidase H from S. plicatus (3). The finding of a varied amino-end for the HPH polypeptide may correlate with the presence of a proximal hydrophobic core ($ser\text{-}leu_4$). Although the protein sequence at the amino terminal of HPH does not meet the requirements of a prokaryotic leader sequence (30), this hydrophobic core might represent a membrane anchorage signal.

At least two other Streptomyces genes are known to initiate with a GTG triplet encoding formyl-methionine (Prof. D.A. Hopwood, personal communication). Therefore, given the high G+C content present in Streptomyces DNA, it does not seem unreasonable to predict that this initiator codon will be found much more frequently in these organisms than in E. coli. We have found that 5 bp preceding the initiator codon, there is a RBS which can interact with the 3'OH terminal of the 16S rRNA from S. lividans (6) with a free energy of -11.6 Kcal/mol (29). This value is similar to those obtained from the binding of the RBS of several genes from Streptomyces starting translation with an AUG codon, such as vph, tsr, aph and pac (2) (4) (J. Vara, R. Lacalle and A. Jiménez, unpublished results). In general, it seems that the RBS from Streptomyces are more similar to those from E. coli (4) than to the RBS from Bacillus subtilis (31). However, it is probable that the RBS from Streptomyces present much closer characteristics between themselves than their counterparts from E. coli.

The initiation of transcription of the hyg gene takes place at/or immediatly next to the guanine at position 157. Therefore, the RNA polymerase recognition sequence(s) must lie upstream from this site. We have not unequivocally demonstrated that the hyg promoter has been cloned. However, there are a number of facts suggesting that we have done so: i) The 3.1 Kb PstI fragment from

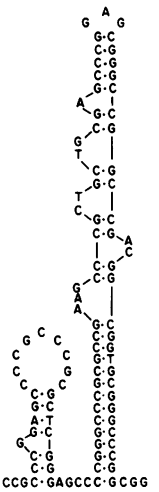

Fig 6. Regions of secondary structure within the hyg gene 3' to the termination of transcription. These structures represent the most stable theoretical conformations (29).

pFM4 has been inserted, in the two possible orientations, in the PstI site of pIJ350 and the two constructions expressed HPH activity in S. lividans (9); ii) The 1.65 Kb PstI-BamHI fragment from pFM4 inserted in a derivative of pIJ922 expressed hygromycin B resistance in S. lividans (10). An examination of the sequence within the MboI insertion site and the PstI site of the aph gene does not reveal any similarity to a promoter region. iii) The structural gene from human interferon $\alpha 2$ has been located in front of the 207 bp NarI-AvaII fragment (Fig 4) and then inserted, in the two possible orientations, in the pIJ702 Streptomyces plasmid. The resulting constructions expressed interferon activity in S. lividans (D. Pulido and A. Jiménez, unpublished observations). When this sequence is compared with those from other cloned Streptomyces promoters (4) (5) weak homology is found in the regions adjacent to the -8 and -30 nucleotides of the vph promoter. No homology to the -10 and -35 consensus promoter sequences of E. coli is present in the hyg promoter, which may explain why this region does not function as a promoter in this bacterium (M. Zalacain, F. Malpartida and A. Jiménez, unpublished observations). It has been suggested that the remarkable ability of

Streptomyces species to transcribe genes from different origins may reside in that these organisms may contain a large number of RNA polymerases with a variety of holoenzyme forms, which are used to activate different sets of genes during their complex life cycles (32). This proposal is in agreement with the finding that at least three types of promoters are present in Streptomyces (4) (33) (5). Whether the hyg promoter can be included within the vph group of promoters or if it belongs to a fourth class is not clear yet.

Termination of transcription in the hyg gene occurs at/or close to the G of position 1252. This region is inmediatly upstream of a small loop (Fig 6) which is also preceding a strong loop. Similar structures have been found in all Streptomyces genes sequenced so far and were assumed to be signals for transcription termination (2) (4). Our results indicate that this may, indeed, be the case.

The biased codon usage in the hyg gene agrees well with the data from the other sequenced Streptomyces genes. However, it is worth noting that while the overall percent of G+C in 3' position is 91%, this value falls to 72% in the first 43 codons, or even more, 10 out of the first 17 codons end by A or T. A similar situation is found in other Streptomyces genes. Thus, 81% of the first 40 triplets in the endo-ß-N-acetylglucosaminidase H gene end by G or C, while this percentage is 93.4% for the rest of the gene (3); and in the ORF-438 sequence preceding the tyrosinase gene (34) these values are 80% for the first 30 triplets and 91.4% for the remaining codons. The meaning, if any, of these variations is not known. It has been suggested that although the Streptomyces tRNA pool can translate all 61 possible codons, their relative abundance may regulate translation of mRNAs with codon usage different to the extremly biased usage of Streptomyces mRNAs. If this holds true, translation of the hyg mRNA might well be regulated by the speed of translation of its initial codons.

Neither the nucleotide sequence or its derived amino acid sequence of the hyg gene shows any degree of homology with the hyg gene present in some Enterobacteriaceae plasmids (35) (36). This finding is not surprising, since the Gram-negative HPH enzy-

mes modify a different OH group of hygromycin B than the S. hygroscopicus enzyme (36) (27). Although both enzymes recognise ATP and hygromycin B it appears that they have separate origins.

ACKNOWLEDGEMENTS

We wish to thank Prof. Julian E. Davies for his interest and comments during the course of this work and for critically correcting the manuscript. We thank Dr. Paul M. Lizardi for his help in the computer analysis of the sequence. We are grateful to Mrs. Asunción Martín for expert technical assistance. This research was supported by grants from the CAICYT (n° 613/165) and by an Institutional grant from the Fondo de Investigaciones Sanitarias to the Centro de Biología Molecular.

+Present address: John Innes Institute, Norwich, UK
§To whom correspondence should be addressed

REFERENCES

1. Chater, K.F. (1984) In Microbial Developments, Losick, R., Shapiro, L. Eds, pp.89-116, Cold Spring Harbor Laboratory, New York.
2. Thompson, C.J. and Gray, G.S. (1983) Proc. Natl. Acad. Sci. USA 80, 5190-5194.
3. Robbins, P.W., Trimble, R.B., Wirth, D.F., Hering, C., Maley, F., Maley, G.F., Das, R., Gibson, B.W., Royal, N. and Biemann, K. (1984) J. Biol. Chem. 259, 7577-7583.
4. Bibb, M.J., Bibb, M.J., Ward, J.M. and Cohen, S.N. (1985) Mol. Gen. Genet. 199, 26-36.
5. Janssen, G.R., Bibb, M.J., Smith. C.P., Ward, J.M., Kieser, T. and Bibb, M.J. (1985) In Genetics of Industrial Microorganisms. Schlesinger, D. Ed, Microbiology, AMS, Washington, DC, in press.
6. Bibb, M.J. and Cohen, S.N. (1982) Mol. Gen. Genet. 187, 265-277.
7. Jaurin, B. and Cohen, S.N. (1984) Gene 28, 83-91.
8. Benveniste, R. and Davies, J. (1973) Proc. Natl. Acad. Sci. USA 70, 2276-2280.
9. Malpartida, F., Zalacain, M., Jiménez, A. and Davies, J. (1983) Biochem. Biophys. Res. Comm. 117, 6-12.
10. Lydiate, D., Malpartida, F. and Hopwood, D.A. (1985) Gene 35, 223-235.
11. Vieira, J. and Messing, J. (1982) Gene 19, 259-268.
12. Thompson, C.J., Kieser, T., Ward, J.M. and Hopwood, D.A. (1982) Gene 20, 51-62.
13. Bibb, M.J., Freeman, R.F. and Hopwood, D.A. (1977) Mol. Gen. Genet. 154, 155-166.
14. Thompson, C.J., Ward, J.M. and Hopwood, D.A. (1980) Nature

286, 525-527.
15. Kieser, T., Hopwood, D.A., Wright, H.M. and Thompson, C.J. (1982) Mol. Gen. Genet. 185, 223-238.
16. Miller, J.H. (1972) In Experiments in Molecular Genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
17. Kieser, T. (1984) Plasmid 12, 19-36.
18. Birnboim, H.C. and Doly, J. (1979) Nucleic Acids Res. 7, 1513-1523.
19. Langridge, J., Langridge, P. and Bergquist, P.L. (1980) Anal. Biochem. 103, 264-271.
20. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) In Molecular Cloning. A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
21. Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol. 65, 499-553.
22. Ambartsumyan, N.S. and Mazo, A.M. (1980) FEBS Letters 114, 265-272.
23. Favaloro, J., Treisman, R. and Kamen, R. (1980) Methods Enzymol. 65, 718-749.
24. Hunkapiller, M.W., Lujan, E., Ostrander, F. and Hood, L.E. (1983) Methods Enzymol. 91, 227-236
25. Hewick, R.M., Hunkapiller, M.W., Hood, L.E. and Dreyer, W.J. (1981) J. Biol. Chem. 256, 7990-7997.
26. Hunkapiller, M.W. and Hood, L.E. (1983) Methods Enzymol. 91, 486-493.
27. Pardo, J.M., Malpartida, F., Rico, M. and Jiménez, A. (1985) J. Gen. Microbiol. 131, 1289-1298.
28. Enquist, L.W. and Bradley, S.G. (1971) Dev. Ind. Microbiol. 12, 225-236.
29. Tinoco, jun., I., Borer, P.N., Dengler, B., Levine, M.D., Uhlenbeck, O.C., Crothers, D.M. and Gralla, J. (1973) Nature New Biol. 246, 40-41.
30. von Heijne, G. (1984) EMBO J. 3, 2315-2318.
31. McLaughlin, J.R., Murray, C.L. and Rabinowitz, J.C. (1981) J. Biol. Chem. 256, 11283-11291.
32. Westpheling, J., Ranes, M. and Losick, R. (1985) Nature 313, 22-27.
33. Hopwood, D.A., Bibb, M.J., Chater, K.F., Janssen, G.R., Malpartida, F. and Smith, C.P. (In press). In Regulation of Gene Expression - 25 years on. Symp. Soc. Gen. Microbiol, Ed, I Booth of C. F. Higgins, Cambridge University Press.
34. Bernan, V., Filpula, D., Herber, W., Bibb, M. and Katz, E. (1985) Gene 37, 101-110.
35. Gritz, L. and Davies, J. (1983) Gene 25, 179-188.
36. Rao, R.N., Allen, N.E., Hobbs, Jr.,J.N., Alborn, Jr, W.E., Kirst, H.A. and Paschal, J.W. (1983) Antimicrob. Agents. and Chemother. 24, 689-695.